

**VISION-GUIDED GRIPPING
OF A CYLINDER**

NAGW- 1333

IN-63-CR

161924 P.12

by

Keith E. Nicewarner and Robert B. Kelley

Rensselaer Polytechnic Institute
Electrical, Computer, and Systems Engineering
Troy, New York 12180-3590

CIRSSE REPORT #93

Vision-Guided Gripping of a Cylinder

KEITH E. NICEWARNER and ROBERT B. KELLEY

Center for Intelligent Robotic Systems for Space Exploration
Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute, Troy, NY 12180

Abstract—The motivation for vision-guided servoing is taken from tasks in automated or telerobotic space assembly and construction. Vision-guided servoing requires the ability to perform rapid pose estimates and provide predictive feature tracking. Monocular information from a gripper-mounted camera is used to servo the gripper to grasp a cylinder. The procedure is divided into recognition and servo phases. The recognition stage verifies the presence of a cylinder in the camera field of view. Then an initial pose estimate is computed and uncluttered scan regions are selected. The servo phase processes only the selected scan regions of the image. Given the knowledge, from the recognition phase, that there is a cylinder in the image and knowing the radius of the cylinder, 4 of the 6 pose parameters can be estimated with minimal computation. The relative motion of the cylinder is obtained by using the current pose and prior pose estimates. The motion information is then used to generate a predictive feature-based trajectory for the path of the gripper.

I. INTRODUCTION

The motivation for the vision-guided gripping of a cylinder is taken from an automated or telerobotic space-assembly scenario. The cylinders in this case are struts similar to those proposed for use in constructing truss structures for the NASA space station, such as depicted in Figure 1.

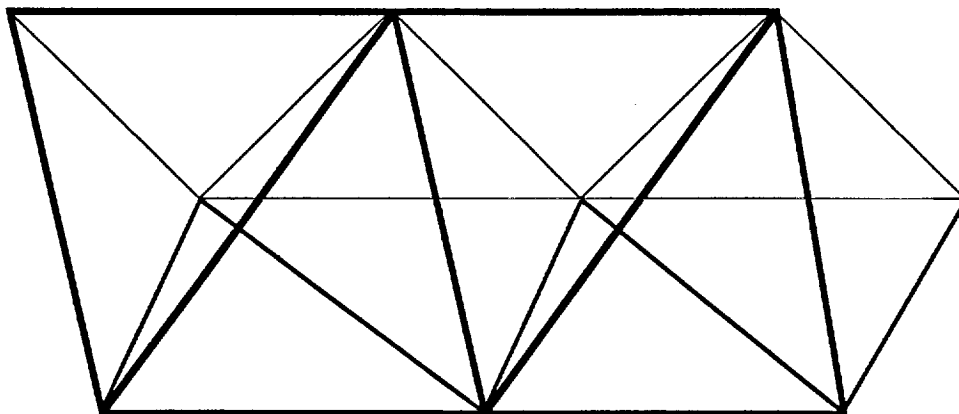


Figure 1. Typical tetrahedral truss structure section.

A typical task for the robot is to insert a strut into a partially constructed truss structure. First, it must locate the strut, given rough estimates of the position of the strut. Once the strut is located,

the robot must visually guide itself to grasp the strut. The strut must then be positioned and oriented for a lateral insertion into cylindrical connector attached to the nodes of the truss structure. The vision system described here can be used not only to guide the grasping, but it can also be used to aid in the insertion of the strut into the connector. This can be accomplished by using two cameras, as indicated in Figure 2. One camera images the strut and the other the connector. Vision can then be used to provide the guidance to the robot to align the strut with the connector.

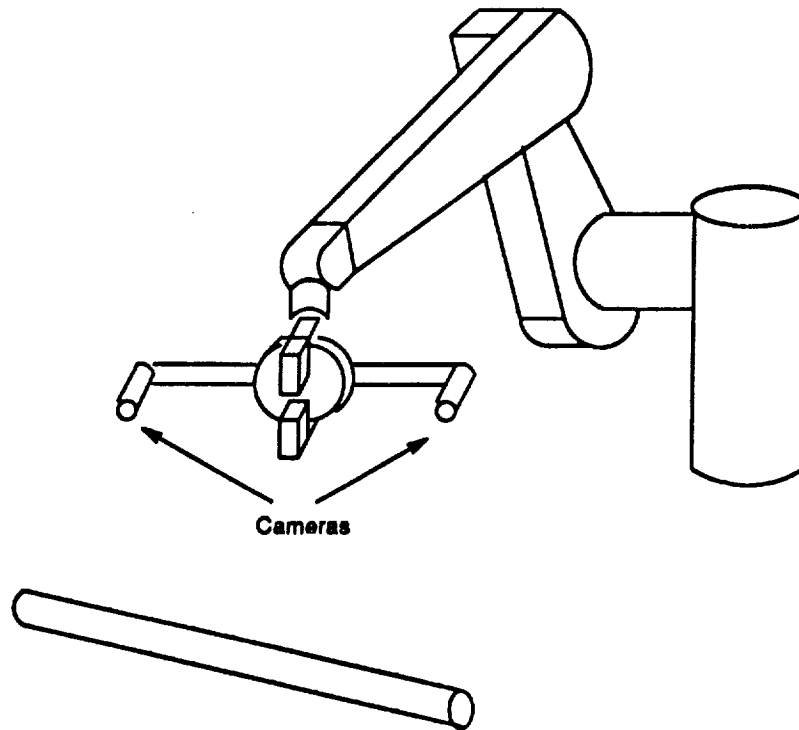


Figure 2. Dual cameras mounted on the gripper.

To perform any type of vision-guided servoing, we need the ability to process the vision information as rapidly as possible. An inherent problem in implementing visual guidance is that conventional vision systems usually have processing cycles that are in the range of hundreds of milliseconds since they are usually tied to the camera frame rate of 33.3 ms (30 Hz) or 40 ms (25 Hz). Robot control systems, on the other hand, usually run with update intervals in the millisecond range. Even if the vision information can be processed instantaneously, the frame rate update interval for standard cameras is still the limiting factor. Thus, to provide accurate and useful information to the robot motion control system, the vision processing system must be able not only to interpolate between frames but also to anticipate the feature motion until the next frame is acquired. To achieve rapid image processing, therefore, it is necessary to limit to the minimum needed for the task the size of the processing region of the image and the amount of processing applied to it. These two ideas, feature extrapolation and minimal information, are essential to realizing real-time, vision-guided servoing.

In this paper we concentrate on a particular instance of vision-guided servoing: directing a robot to grasp a cylinder. Monocular information from a gripper-mounted camera is used to servo the gripper to grasp a cylinder. As compared to stereo methods which employ two cameras, using only one camera greatly simplifies the image processing requirements. Three-dimensional information

can be obtained from the motion of the robot. Thus, the need for rapid pose estimates of the cylinder is clear.

There are many image processing techniques that give the position and orientation of various objects. Some methods are developed to handle as many sizes and shapes as possible, thereby sacrificing speed for versatility. Other methods are developed for the purpose of processing only limited types and numbers of objects, thereby increasing the speed of the image processing. There are a few methods that deal specifically with cylinders [1, 2, 3]. These methods are based on performing some sort of inverse perspective transformation on the image to obtain an anticipated object surface. This transformed surface is then matched with the surface of the target object. Surface matching techniques have not yet demonstrated the speed needed for vision-guided servoing.

Once the image processing method has been established, a frame extrapolation method and predictive feature tracking algorithm must be specified. We want to align the gripper jaws with the axis of the cylinder in a continuous error-reducing fashion in the manner of [4, 5], as opposed to the traditional look-and-move approach [6].

II. CYLINDER POSE ESTIMATION

A. Recognition Phase

Before we can begin to rapidly estimate cylinder pose, there are several presumed conditions that must be verified. These are:

1. There is a cylinder in the field of view,
2. There is a sufficient length of the cylinder visible in the image, and
3. The edges of the cylinder can be extracted.

The first condition seems trivial. However, it is critical to establish whether there is anything in the image on which to perform vision-guided servoing. It is important to verify that the object is cylinder-like. The second condition requires that the cylindrical segment is at least twice as long as it is wide. This assures that accurate pose estimates can be obtained. These first two conditions are verified by first looking at a thresholded image and computing the moments of inertia of any blobs, and looking for any that are "long and thin."

The third condition is checked by applying several edge-detection methods and choosing the most appropriate one. The third condition considers first a simple threshold edge detection scheme to check if the cylinder has sufficient contrast with the background. If there is insufficient contrast resulting in too much background clutter, then a gradient operator is used to detect high spatial-frequency peaks. Distant clutter is usually out of focus and, thereby, not a problem since the camera is focused at a distance just beyond the end of the gripper.

Once the three conditions have been verified, the scan lines are chosen and the radius of the cylinder is estimated, if it is not known *a priori*. The scan line positions and the scan ranges, which define the scan region, are chosen to minimize the noise that will be encountered during the servo phase. The scan lines are chosen to be as far apart as possible to assure more accurate pose estimates. The pose estimation algorithm works fastest with scan lines that are aligned with either rows or columns. Hence, if the axis of the cylinder is more horizontal than vertical, vertical scan lines (columns) are used.

B. Servo Phase

The servo phase begins by processing the scan lines for edges, or *critical points*. If some unexpected noise is encountered while scanning for critical points, the scan ranges or the scan line positions can be adjusted. Later it will be shown that 4 of the 6 cylinder pose parameters can be determined from only 4 critical points. The pose of the cylinder is computed relative to the camera. The 4 pose parameters are:

1. θ , the clockwise rotation of the cylinder about the optical axis, relative to the image plane vertical;
2. x_C , the horizontal displacement of the center of the cylinder from a vertical plane through the optical axis;
3. ϕ , the tilt angle of the cylinder axis out of a plane normal to the optical axis; and
4. d_C , the distance from the camera to the center of the cylinder on the optical axis.

The servo target is to position and orient the image of the cylinder so the cylinder axis is aligned vertically and passes through the center of the image plane. The servo target provides the robot control system with an error measure that can be driven to zero.

III. DERIVATION OF THE POSE PARAMETER ESTIMATES

Consider a $W \times H$ grey-scale image of a scene containing a cylinder as in Figure 3. For this discussion, it is assumed that two critical points lie on the top raster line, the top row, and the other two critical points lie on the bottom raster line, the bottom row. The cylinder is oriented vertically and the critical points are designated as x_{TL} , x_{TR} , x_{BL} , and x_{BR} .

Using the critical points, the orientation of the symmetry axis of the cylinder is computed using geometry. If the critical points are less than 10° from the optical axis, which passes through the center of the image plane $(0, 0)$, then the perspective distortion effects are negligible. Under these conditions, the centerline of the cylinder image is a good approximation to the location of the symmetry axis of the cylinder.

The centerline and its rotation from the vertical are computed as follows. Define the midpoints of the top and bottom critical point pairs x_{TC} and x_{BC} . Thus,

$$x_{TC} = \frac{1}{2}(x_{TL} + x_{TR})$$

and

$$x_{BC} = \frac{1}{2}(x_{BL} + x_{BR}) .$$

Then the clockwise rotation of the cylinder θ about the optical axis, relative to the vertical, is given by

$$\theta = \tan^{-1} \left(\frac{x_{TC} - x_{BC}}{H} \right) .$$

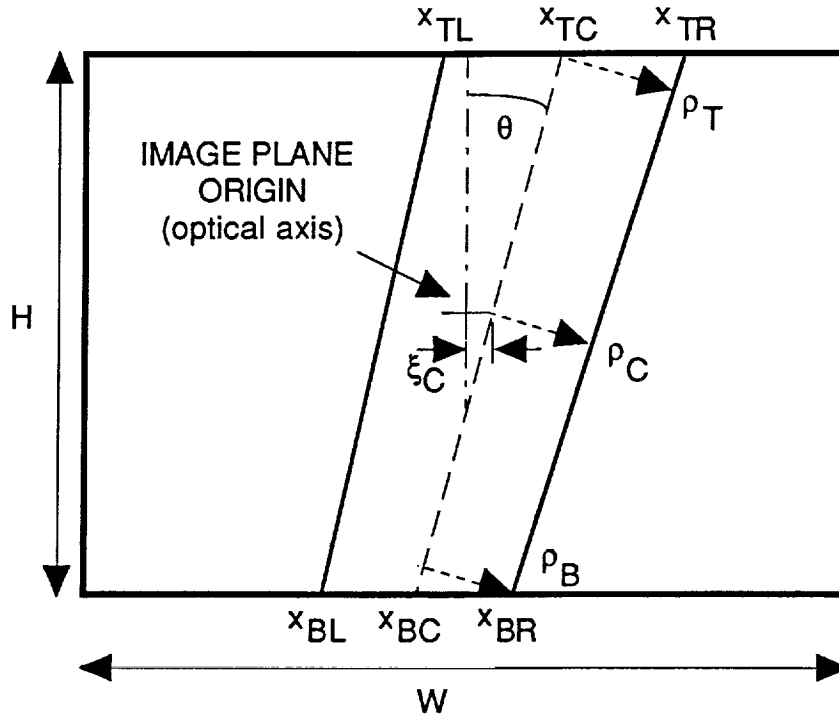


Figure 3. Image plane geometry for a cylinder.

Within the image, the apparent center of the cylinder has an x-component ξ_C given by the midpoint between the top and bottom center points. Therefore, the apparent horizontal displacement of the center of the cylinder from the optical center is

$$\xi_C = \frac{x_{TC} + x_{BC}}{2} .$$

The apparent radius of the cylinder at the top of the image is given by

$$\rho_T = (x_{TR} - x_{TC}) \cos \theta ,$$

and, similarly at the bottom of the image,

$$\rho_B = (x_{BR} - x_{BC}) \cos \theta .$$

These parameters are used to calculate the apparent radius in the center of the image,

$$\rho_C = \frac{l}{2} (\rho_T + \rho_B).$$

The tilt angle ϕ is the angle the cylinder axis is rotated towards the viewer relative to a plane normal to the optical axis. The tilt angle geometry is shown in Figure 4.

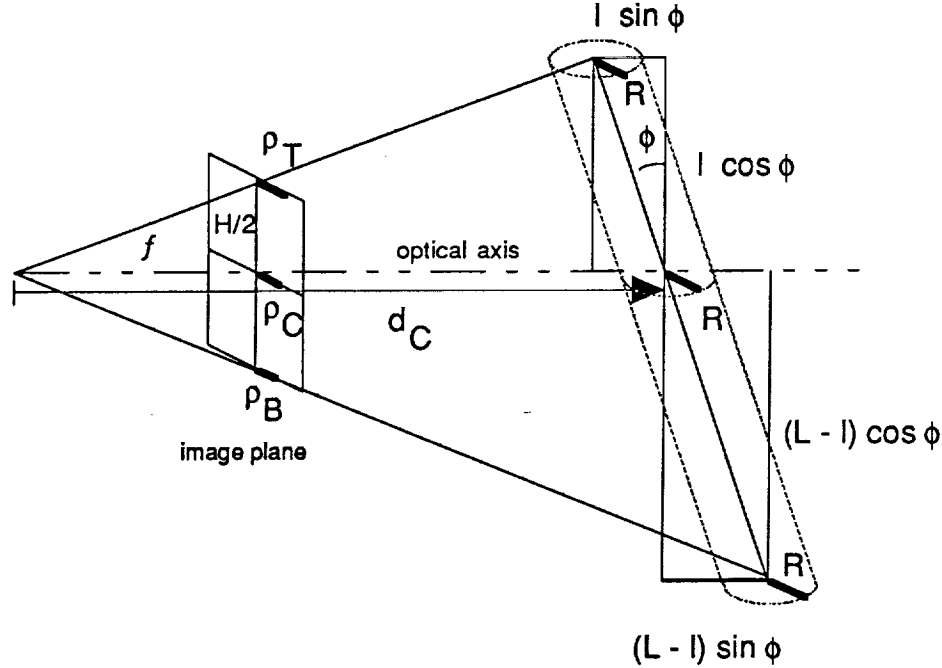


Figure 4. Tilt angle geometry.

Assuming a constant optical magnification factor μ (i.e., $d_C \gg f$), the following expressions are obtained:

$$\mu = \frac{R}{\rho_T} = \frac{d_C - l \sin \phi}{f}, \quad \mu = \frac{R}{\rho_B} = \frac{d_C + (L - l) \sin \phi}{f} \quad \text{and} \quad \mu = \frac{R}{\rho_C} = \frac{d_C}{f},$$

where R is the actual radius of the cylinder, d_C is the distance to the center of the cylinder, L is the length of the visible portion of the cylinder, l is the length above the optical axis, and f is the focal length of the camera. (Note, the focal length must be expressed in the same units as the distance and length measurements.) The terms d_C , L and l are not known *a priori*. However, they do not appear in the final expression for the tilt angle ϕ given below.

Using similar triangles, we get a second set of identities:

$$\frac{H}{2f} = \frac{l \cos \phi}{\frac{\rho_T R}{f}} = \frac{(L - l) \cos \phi}{\frac{\rho_B R}{f}},$$

where H is the vertical distance between the scan lines.

Simplifying and solving for ϕ , the tilt angle is obtained as

$$\phi = \tan^{-1} \left(\frac{2f}{H} \left(\frac{\rho_T - \rho_B}{\rho_T + \rho_B} \right) \right).$$

The distance to the observed center of the cylinder is then given by

$$d_C = \frac{R f}{\rho_C},$$

where R is the actual radius of the cylinder.

The angle in the image plane of the cylinder axis relative to the vertical is given by

$$\psi = \tan^{-1} \left(\frac{\xi_C}{f} \right),$$

where f is the focal length of the camera.

And finally, the lateral displacement x_C is given by

$$x_C = d_C \sin \psi.$$

The orientation angles θ and ϕ can be determined from these expressions without a knowledge of the actual radius of the cylinder. However, the position parameters of the cylinder d_C and x_C are expressed in terms of the actual radius of the cylinder. Therefore, with two parallel scan lines from the image and *a priori* knowledge of the actual radius of the cylinder, 4 of the 6 pose parameters of the cylinder can be determined.

IV. PREDICTIVE FEATURE TRACKING

The vision updates are being calculated every 33.3 ms in the ideal case. The actual time may vary if, for example, the cylinder is lost and must be re-acquired. The robot controller runs at a fixed update rate, 4.5 ms in our case. Thus we need to interface the asynchronous vision updates with the controller. By taking the current pose along with several prior poses, a linear approximation to the relative motion of the cylinder can be constructed. Using this approximation, we generate set-points every 4.5 ms along an anticipated path for the robot. This works well if we assure the vision updates will be regular. If the cylinder is lost, the robot will follow the approximated path indefinitely. However, if we place an additional criterion on the approximation that the velocity

along the trajectory should go to zero after, for example, 66.7 ms, the robot motion halts gracefully while the cylinder is re-acquired.

V. EXPERIMENTAL RESULTS

The experimental set-up consists of a compact camera mounted on the gripper of a PUMA robot arm and a set of white cylinders of various diameters (8 mm, 16 mm, 22 mm, and 38 mm). The camera has a 24 mm focal length and an imaging plane of 570×485 pixels or 6.39×4.88 mm. The program is written in C under a VxWorks real-time environment. A Datacube vision system performs the image acquisition and processing. Another program interfaces the PUMA's controller to run under VxWorks.

The edge detection method employed is thresholding and is performed by the Datacube at frame rate. The program calculates the pose parameters and sends them to the robot controller program via internet sockets. The robot interface program gives position updates to the PUMA every 4.5 ms. This process is continued until the user exits the program. If the cylinder is lost at anytime, the scan lines are adjusted until the cylinder is re-acquired. If no cylinder is found in the image, the robot halts and the program waits until one appears.

Approximately 70 trials were made over a variety of conditions. These final values are obtained by using the robot and its own calibration to measure the errors. The robot is positioned and a pose estimate is computed. Then the robot is moved slightly and another pose estimate computed. By comparing the visual change with the change in the position of the robot, the error is computed. The results from the cylinder pose estimation trials and the ranges of the parameter values are shown in Table 1.

Table 1. Pose Estimation Accuracy

Parameter	Range	Average Absolute Error
θ	0 - 90°	0.28°
ϕ	0 - 45°	3.15°
d_C	13 - 46 cm	1.20 mm
x_C	0 - 10 mm	0.13 mm

The best accuracy for θ is obtained for θ is zero and for ϕ when both θ and ϕ are zero. The best accuracy for d_C is obtained at a distance of about 15 cm and for x_C at 15 cm when x_C is zero.

Used in conjunction with an instrumented gripper with V-fixture jaws, such as shown in Figure 5, this visual guidance technique provides adequate accuracy to permit cylinders with significant initial pose uncertainties to be grasped efficiently.

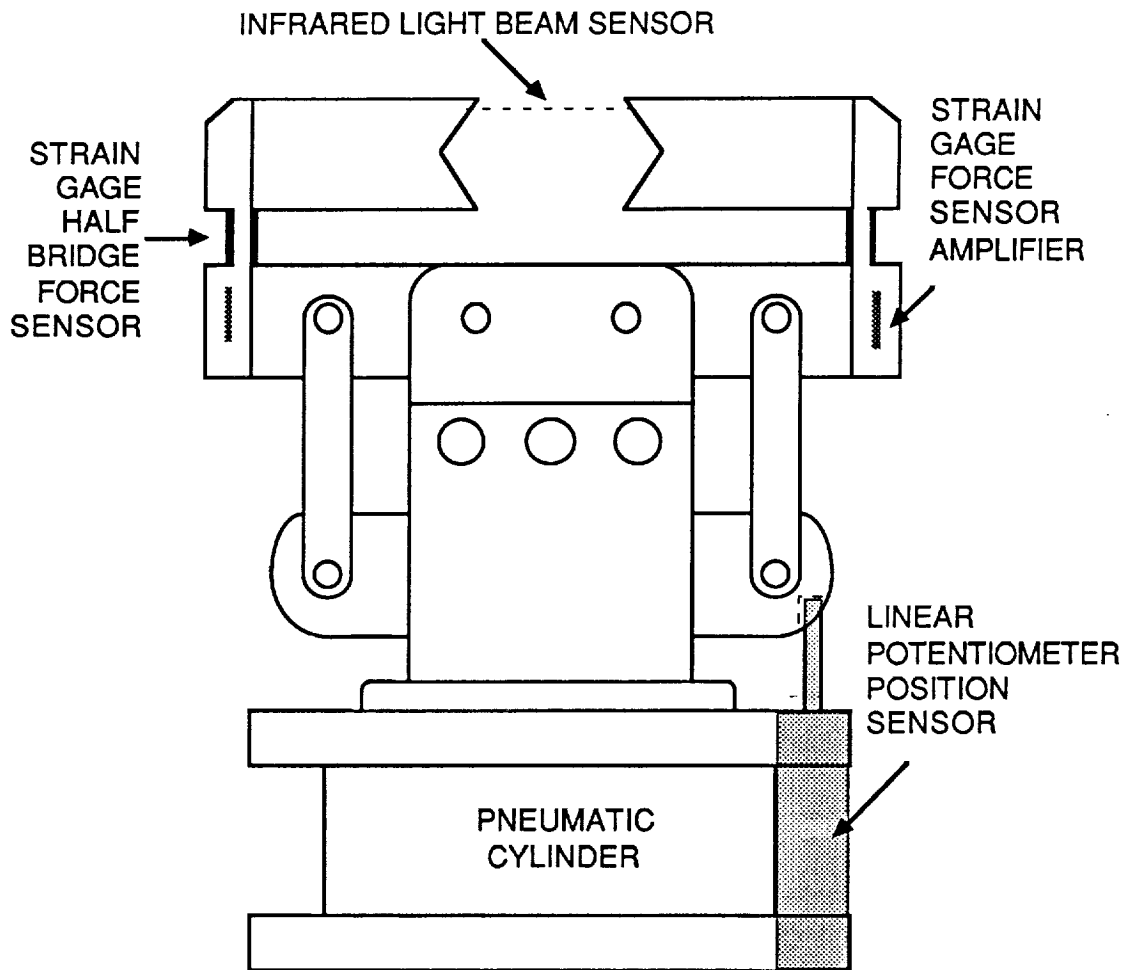


Figure 5. CIRSSSE general purpose, instrumented, pneumatic powered gripper.

VI. SUMMARY

A frame-rate method for cylinder pose estimation has been presented. The algorithm estimates the pose relative to the camera using a single camera mounted on the gripper of a robot. Frame-rate updates are calculated to align the gripper with the cylinder so that, for example, a minimum reaction force and torque is results when the cylinder is grasped. The method scans two parallel, widely spaced lines in the image. These scan lines are then processed for high contrast edges, called critical points. The four critical points represent the edges of the cylinder in the image. From these points, the three-dimensional orientation of the cylinder is estimated. The position of the cylinder is estimated using the camera parameters, *a priori* knowledge of the diameter of the cylinder, and the four critical points. This pose estimation method is combined with a simple predictive feature extrapolation technique to provide inputs into a real-time vision-guided servoing process.

ACKNOWLEDGEMENT

This research is performed at the Center for Intelligent Robotic Systems for Space Exploration at Rensselaer Polytechnic Institute and is supported in part by Grant number NAGW-1333 from NASA.

REFERENCES

- [1] H. Prinz, "Finding the orientation of a cone or a cylinder," *Proc. Workshop on Computer Vision*, IEEE Computer Society Press, 1987, pp. 94-99.
- [2] M. Richetin, M. Dhome and J.T. Lapreste, "Inverse perspective transformation from zero-curvature curve points, application to the localization of some generalized cylinders," *Proc. Conf. on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1989, pp. 517-522.
- [3] M.J. Korsten and Z. Houkes, "Parametric descriptions and estimation, a synergetic approach to resolving shape from shading and motion," *Proc. Third Int. Conf. on Image Processing and its Applications*, IEE, 1989, pp. 5--9.
- [4] J.T. Fedemma and O.R. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Trans. on Robotics and Automation*, 5(5), 1989, pp. 691-700.
- [5] J.T. Fedemma and C.S.G. Lee, "Adaptive image feature prediction and control for visual tracking with a hand-eye coordinated camera," *IEEE Trans. on Systems, Man and Cybernetics*, 20(5), 1990, pp. 1172-1183.
- [6] Y. Shirai and H. Inoue, "Guiding a robot by visual feedback in assembly tasks," *Pattern Recognition*, 5, 1973, pp. 99-108.